International Journal of Informatics and Communication Technology (IJ-ICT)
Vol. 15, No. 2, June 2026, pp. 456~464
ISSN: 2252-8776, DOI: 10.11591/ijict.v15i2.pp456-464

A critical review of information retrieval techniques: current trends and challenges

Sanket D. Patil[1,2], Zahir Aalam[1]
[1]Department of Information Technology, Thakur College of Engineering and Technology, University of Mumbai, Mumbai, India
[2]Department of Computer Engineering, Vidyavardhini's College of Engineering and Technology, University of Mumbai, Mumbai, India

Article Info

Article history:

Received Oct 15, 2024
Revised Nov 6, 2025
Accepted Dec 14, 2025

Keywords:

Deep learning
Information retrieval
IR system
LSTM
NLP

ABSTRACT

The realm of information retrieval is witnessing transformative advancements, driven by the integration of deep learning techniques, specialized algorithms, and domain-specific applications. Information retrieval systems play an important role in many applications including in the Artificial Intelligence powered systems that can be seen in many applications. Information Retrieval, generally, acts an important task in the knowledge discovery phase of any query based intelligent system. This paper presents a comprehensive review by conducting a detailed analysis of the technological nuances, dataset specifications, and pivotal findings. This detailed review has been done with the special emphasis on the kind of technology used to achieve accurate information retrieval, domain of the study, and the system's ability to retain or work with tables and figures, among other parameters. Navigating through the rich tapestry of methodologies, the paper underscores the pivotal role of deep learning frameworks in revolutionizing traditional retrieval paradigms. Furthermore, it sheds light on the innovative integration of textual information, algorithmic advancements, and specialized datasets to enhance the efficacy and granularity of information retrieval mechanisms.

Corresponding Author:

Sanket D. Patil
Department of Information Technology, Thakur College of Engineering and Technology
University of Mumbai
Mahatma Gandhi Road, Mumbai, India
Email: sanketpatilvcet@gmail.com

1. INTRODUCTION

In an age defined by digital proliferation, the intricacies of information retrieval have become both an academic pursuit and an operational imperative. The intriguing dance between human cognition, computational algorithms, and vast datasets has paved the way for ground-breaking research in many fields of research and businesses. The 'data' has been at the heart of many artificial intelligence/machine learning/ deep learning breakthroughs [1].

Retrieval and extraction of relevant information from the humongous repository of data is an important task. This task involves streamlining the process of extracting meaningful insights from voluminous information repositories. An IR system can be defined as a software application that manages the organisation, storing, retrieving, and assessing information from document repositories, especially textual information, which is known as information retrieval (IR). Information retrieval is the process of getting content from vast computer-stored collections that may typically be recorded in an unstructured manner, i.e., text, which meets a demand for information [2].

The uses of IR systems are manifold [3]:

1. Web search engines: Platforms like Google, and Bing employ advanced IR techniques to index and retrieve web pages based on user queries.
2. Digital libraries: IR systems facilitate the organization, storage, and retrieval of digital resources in libraries and academic repositories.
3. E-commerce: Product search functionalities on online shopping platforms rely on IR systems to display relevant products to users.
4. Healthcare: IR plays a pivotal role in medical databases, aiding professionals in accessing relevant research papers, patient records, and diagnostic information.

The importance of the IR systems is crucial in many applications and usecases across sectors [4]:

1. Efficient access to information: Information retrieval (IR) systems enable users to swiftly access and retrieve relevant information from vast repositories, thereby optimizing time and effort.
2. Decision making: In various sectors, from healthcare to business, timely and accurate information retrieval facilitates informed decision-making processes.
3. Knowledge discovery: IR systems can aid researchers and professionals in discovering new insights, trends, and patterns within large datasets, fostering innovation and discovery.
4. Enhanced user experience: In the digital age, user experience hinges on the seamless retrieval of information, making IR systems indispensable for platforms ranging from search engines to e-commerce sites.

There are many applications of information retrieval [4]:

1. Personalized recommendation systems: Platforms like Netflix and Amazon utilize IR techniques to offer personalized content and product recommendations to users.
2. Legal and regulatory compliance: In the legal sector, IR systems aid professionals in retrieving pertinent case laws, statutes, and regulatory guidelines.
3. Content management systems: IR plays a crucial role in content management platforms, facilitating the efficient organization, retrieval, and dissemination of digital content.
4. Social media analysis: IR techniques are employed to analyze and retrieve relevant information from social media platforms, aiding businesses in market analysis, sentiment analysis, and trend forecasting.

Having established the importance of Information Retrieval, it is also important to note that there are many challenges that make the process of information retrieval a big challenge [5]:

1. Information overload: The exponential growth of digital data has exacerbated the challenge of information overload, necessitating more refined retrieval techniques.
2. Ambiguity and context sensitivity: Language is inherently ambiguous, and IR systems must grapple with the nuanced meanings and contextual variations of user queries.
3. Dynamic nature of data: The real-time and dynamic nature of many datasets poses challenges in ensuring the relevancy and accuracy of retrieved information.
4. Algorithmic Biases: IR systems can inadvertently perpetuate biases present in training data, leading to skewed or unfair information retrieval outcomes.

In summary, information retrieval stands as a cornerstone in the digital ecosystem, underpinning a myriad of applications and domains [6], [7]. While it offers unparalleled benefits in facilitating access to information and driving innovation, the field also presents intricate challenges that necessitate continuous research, refinement, and ethical considerations [8], [9].

By delving deep into the methodologies, datasets, and the status of retention of tables and pictures from the original text and findings of each paper, this paper aspires to achieve multiple objectives. Firstly, it aims to provide a granular understanding of the technological foundations that underpin contemporary information retrieval systems. Secondly, it seeks to elucidate the domain-specific applications and challenges, offering readers a nuanced perspective on the practical implications of these research endeavours.

Furthermore, by synthesizing insights from these diverse studies, this survey endeavours to identify emergent trends, potential research gaps, and avenues for future exploration in the field of information retrieval.

2. INFORMATION RETRIEVAL SYSTEMS

The issue that Pandey and Bhat [10] identified in natural language processing, text-based question answering (QA) is a crucial activity that seeks to deliver precise and pertinent responses to users' inquiries. Conventional methods of quality assurance depended on manually designed features and rule-based systems. On the other hand, current developments in deep learning and information retrieval have demonstrated encouraging outcomes in enhancing the functionality of QA systems. The performance of quality assurance has significantly improved as a result of the integration of deep learning techniques with IR. The authors intend to capitalize on the advantages of both IR and deep learning models. A hybrid models have been developed, such as the bi-encoder and tri-encoder architectures.

The authors examine several suggested models and datasets available for the job, assess their effectiveness using performance metrics, and give a brief synopsis of the QA system and its many domains and subdomains in order to study textual question answering. The authors do not divulge whether the proposed system is able to etrieve the images or the tables in the original data while answering to the queries posted by the users. Also, the paper makes use of the existing technologies.

Van *et al.* [11] plans to provide effective deep learning-based techniques in this research for processing legal documents, including tasks related to legal document retrieval and legal question answering in the automated legal question answering competition (ALQAC 2022). The XLM-RoBERTa model, which is pre-trained from a sizable quantity of unlabeled corpus before being fine-tuned to the particular tasks, is the basis of this technique. The procedure performs effectively in legal retrieval information tasks with a little amount of labeled data, according to the experimental results. Furthermore, the authors claim that various information retrieval tasks in low-resource languages may be accomplished using this methodology. Again, the authors do not divulge whether they are considering or working with the images or the tables in the original data while answering to the queries posted by the users.

Zhao *et al.* [12] pointed out that the problem with a recommendation system's overall operation is that it receives candidate items, the retrieval layer filters a small number of items that users might find interesting, and the ranking layer performs more precise filtering to produce the final recommendation list. The retrieval layer is the initial step in the recommendation system, and its effectiveness influences the upper bound of the ranking impact as well as how rapidly the system can filter information. The authors concentrate on deep learning techniques frequently employed in the retrieval layer to address the stated issue, which is relevant to the optimization of recommendation systems.

The paper by Matsiuk *et al.* [13] delves into the application of an information retrieval thesaurus to enhance information retrieval technologies within specific data domains. In the context of the accelerating growth of big data, the authors explore the role of a thesaurus in refining the efficiency and effectiveness of information retrieval processes. The primary objective of the study is to demonstrate the practical implementation of an information retrieval thesaurus in optimizing information retrieval technologies. The focus is on specific data domains, suggesting a targeted approach to addressing challenges related to domain-specific information organization and retrieval.

The authors employ a practical approach, emphasizing the application of an information retrieval thesaurus. The study likely involves the development and implementation of the thesaurus within a defined data domain. The methodology is expected to include processes such as lexical system analysis, terminological data classification, and the integration of the thesaurus into existing information retrieval systems. The paper likely presents findings regarding the effectiveness of the information retrieval thesaurus in improving search processes within specific data domains. This may include insights into enhanced search accuracy, reduced information overload, and the alignment of user queries with document content, leading to more relevant search results.

The primary objective of the Scholtes [14] is to investigate the potential of unsupervised learning methodologies in addressing challenges associated with information retrieval. Given the complexities and intricacies of information retrieval systems, the paper delves into the application of neural network-based unsupervised learning approaches as a means to enhance retrieval accuracy and efficiency.

The paper likely presents findings highlighting the efficacy of unsupervised learning techniques in improving various facets of information retrieval, such as Enhanced retrieval accuracy through neural network-based models, Potential for automated feature extraction and representation learning, and adaptability and scalability of unsupervised learning approaches in diverse retrieval scenarios.

Benedetto *et al.* [15] identify the burgeoning issue of information overload within educational technology systems, wherein learners are inundated with vast amounts of content, often leading to cognitive overwhelm and reduced learning efficacy. The authors highlight the inherent complexities in navigating expansive content repositories, emphasizing the need for innovative solutions to streamline information delivery and promote effective learning.

To address the aforementioned challenges, authors advocate for the integration of advanced summarization techniques within educational platforms. The authors propose the development and implementation of tailored summarization algorithms capable of distilling complex educational content into concise and digestible summaries without compromising essential information. By leveraging these summarization methodologies, the authors contend that educational technology systems can facilitate more efficient content navigation, enhance user engagement, and foster an adaptive learning environment tailored to diverse learner needs.

Aksonov *et al.* [16] highlight the pervasive challenge of information overload, wherein traditional question-answering systems grapple with the intricacies of processing and retrieving relevant information from expansive data repositories. The authors identify inherent scalability constraints within existing

question-answering systems, emphasizing the limitations in efficiently processing increasing data volumes without compromising system performance. Authors elucidate the complexities associated with heterogeneous data sources and formats, underscoring the challenges in harmonizing disparate data elements to facilitate coherent question-answering processes.

The research conducted by the authors offers a seminal contribution to the discourse on question-answering systems and big data analytics. By elucidating the challenges of information overload, system scalability, and data complexity, and proposing innovative solutions grounded in advanced analytics and adaptive algorithmic design, the authors provide a comprehensive framework for harnessing the transformative potential of big data in redefining the capabilities and efficacy of question-answering systems in the digital age.

Xu [17] identifies inherent inefficiencies in traditional information retrieval mechanisms, emphasizing challenges in prioritizing and ranking information based on relevance and significance. The author elucidates the challenges posed by semantic ambiguities and contextual variability within textual data, highlighting the complexities associated with deciphering and interpreting information in diverse contexts.

To address the identified challenges, author advocates for the integration of the TextRank algorithm, a graph-based ranking model inspired by the PageRank algorithm. By elucidating the challenges associated with traditional retrieval mechanisms and proposing an innovative solution grounded in graph-based ranking and semantic analysis, the author provides a comprehensive framework for harnessing the transformative potential of TextRank in redefining the landscape of information retrieval in the digital age.

The problem identified by Kanhaiya *et al.* [18] underscores the multifaceted nature of legal judgments, characterized by intricate legal terminologies, nuanced interpretations, and contextual dependencies. The authors highlight the challenges posed by the inherent complexities of legal language, emphasizing the limitations of traditional information retrieval mechanisms in effectively navigating and extracting pertinent information from judgements The authors identify inherent inefficiencies in existing information retrieval systems within the legal domain and the challenges arising from semantic ambiguities and contextual variability inherent in legal judgements, underscoring the complexities associated with deciphering and interpreting legal language in diverse contexts.

To address the identified challenges, Kanhaiya et al. propose the development of an AI-enabled information retrieval engine (AI-IRE) specifically tailored for legal services. The authors delineate the architecture and functionalities of AI-IRE, emphasizing its capability to integrate expert-annotated natural language processing (NLP) techniques with advanced machine learning algorithms to facilitate efficient and contextually relevant information retrieval from legal judgements.

In the era of big data, the development of robust and scalable information retrieval frameworks is paramount to efficiently navigate and extract meaningful insights from vast and heterogeneous data repositories. Nguyen *et al.* [19] embark on a comprehensive exploration into the intricacies of crafting such frameworks tailored for the big data landscape. Authors identify inherent challenges associated with scalability and performance limitations within existing information retrieval frameworks, particularly in the context of big data. The complexities arising from the heterogeneity and inherent variability of big data emphasize challenges in harmonizing diverse data types, structures, and formats within traditional retrieval frameworks. The challenges posed by ensuring robustness and reliability within big data information retrieval frameworks, particularly in mitigating risks associated with data inconsistencies, redundancies, and inaccuracies.

To address the identified challenges, the authors propose the development of a robust information retrieval framework tailored for big data contexts. The authors advocate for the integration of innovative algorithms designed to address the complexities of heterogeneous data processing within big data environments. The integration of resilience and trustworthiness mechanisms within the proposed framework aims to mitigate risks associated with data inconsistencies and inaccuracies. The authors highlight the implementation of advanced validation, verification, and error-correction mechanisms to foster consistent, accurate, and reliable retrieval outcomes, thereby ensuring the integrity and trustworthiness of information accessed within big data contexts.

In the domain of information retrieval systems, the generation of high-quality, relevant judgments stands as a critical component in evaluating system efficacy and performance. Joseph and Ravana [20] delve into the intricacies of enhancing the evaluation process by leveraging document similarity and document pooling techniques. Authors identify inherent limitations associated with conventional evaluation metrics utilized in assessing information retrieval systems, challenges stemming from ambiguities and inconsistencies prevalent in manual relevance judgments, and the necessity for advancing evaluation methodologies to foster enhanced efficacy, insightfulness, and actionable insights.

To address the identified challenges, the authors advocate for the integration of document similarity techniques within the evaluation framework, the implementation of document pooling techniques aimed at enhancing the generation of relevance judgments, and the deployment of document similarity and pooling techniques is posited to advance existing evaluation paradigms by fostering enhanced granularity, accuracy, and contextuality in relevance judgments.

Srivarsha and Manikandan [21] tackled the challenge of content-based image retrieval (CBIR) by integrating textual information. Recognizing the inherent limitations of conventional CBIR systems that rely solely on visual features, the authors highlighted the potential of textual data in enhancing retrieval accuracy and relevance. A primary problem they addressed was the difficulty users often face when trying to retrieve specific images based on visual content alone, especially when images might be ambiguous or when semantic context is crucial.

To bridge this gap, the authors introduced a novel scheme that harnessed textual information associated with images. By leveraging textual metadata, annotations, or accompanying descriptions, their scheme aimed to provide a richer, contextually relevant set of results. The integration of textual information offered users a more nuanced and precise mechanism to retrieve images, especially in scenarios where visual cues might be insufficient or misleading. Overall, their research underscored the synergistic potential of combining visual and textual data in advancing the capabilities and accuracy of content-based image retrieval systems.

According to Vekariya and Limbasiya [22], the processing of questions and excellent response processing is the main function of the question-answering system. Additionally, the study contends that the precise range of questions that excellent responses provide is the main issue with question answering. Furthermore, by using this, it ought to take less time to find a similar query and an excellent solution. Nonetheless, obtaining accurate data is essential.

Therefore, the research suggests using two distinct information retrieval strategies in order to obtain the appropriate and highly accurate replies. The study suggests using the deep long short-term memory method with versatile global T-max pooling for knowledge retrieval. Using word2vec, the words are first transformed into vectors. After that, information retrieval is accomplished using these two methods. The article then suggests ranking the output from these two systems by using a deep factorization machine. The response that is deemed to be the most efficient is chosen. In order to assess the effectiveness of the suggested approach, the research suggests using three distinct datasets. Before testing the system, the paper performs an analysis of different word embedding techniques on the three datasets identified for experimentation.

The publication states that it has attained very excellent accuracy, with over 80% accuracy. However, the study omits to describe the system architecture in depth. In other words, while the publication notes the usage of deep long short-term memory, it withholds information about the network's precise design, including details about the number of neurons employed and the network's depth. Furthermore, the relative impact of the pooling layer in the final classification is unclear due to the concurrent use of the versatile global T-max pooling and deep long short-term memory approaches. The paper does not discuss whether or not tables and images are retained.

3. RESULTS AND DISCUSSION

Table 1 provides the analysis of all the papers reviewed in the previous section. The said analysis has been conducted on the following parameters: technology used, domain of study, retrieval of original table, retrieval of original pictures, and key findings of the papers. The table provides a comprehensive summary of 18 academic papers, each employing different technologies to enhance information retrieval and related processes across various domains. The key findings are grouped according to the technology used and the specific domain of application.

Firstly, papers utilizing cosine Similarity report significant advancements. In the domain of QA, the integration of information retrieval with deep learning techniques has shown to significantly boost the performance of QA systems. Similarly, in the field of education, the use of summarization techniques has been found to enhance educational technology systems.

Recurrent neural networks (RNN) have been applied in the legal QA domain, where the implementation of deep learning methods has improved the accuracy of legal question-answering systems. This highlights the potential of RNNs in handling complex legal texts and providing precise answers.

Several papers focus on the use of long short-term memory (LSTM) networks. In Recommendation Systems, LSTM has enhanced retrieval algorithms, leading to better recommendations. Within QA systems, big data analysis has facilitated the development of more effective QA frameworks. Furthermore, a proposed framework for scalable information retrieval in big data contexts and another application in content-based image retrieval underscore the versatility and effectiveness of LSTM in various retrieval tasks.

The use of thesaurus generation in search engines has demonstrated improvements in retrieval technologies within specific data domains. This suggests that thesaurus-based approaches can enhance search precision and relevance. Neural networks have been applied to free text database search, where unsupervised learning techniques have addressed key challenges in information retrieval. This indicates the potential for neural networks to manage and interpret vast amounts of unstructured data.

The TextRank algorithm has been employed in search engines, improving information retrieval through advanced ranking algorithms. This shows the efficacy of TextRank in enhancing search engine performance by better ranking relevant documents.

NLP techniques, when annotated by experts, have been shown to significantly enhance AI-enabled information retrieval in legal services. This highlights the importance of expert involvement in refining NLP models for specialized domains.

Finally, document-level similarity techniques in the legal domain have been shown to improve the quality of relevance judgments in information retrieval systems. This involves using document similarity and pooling techniques to better judge the relevance of legal documents. Deep LSTM techniques applied to question answering demonstrate effective information retrieval capabilities, reinforcing the value of deep learning in enhancing QA systems.

Table 1. Overview of existing information retrieval framework, technologies, domain and key findings

Paper	Technology used	Domain of study	Retrieval working with tables	Retrieval/ working with tables	Key findings
[10]	Cosine Similarity	Question answering	No	No	Integration of information retrieval and deep learning enhances textual question answering systems
[11]	RNN	Legal question answering	No	No	Deep learning improves accuracy in legal question answering.
[12]	LSTM	Recommendation system	No	No	Deep learning enhances retrieval algorithms in recommendation systems.
[13]	Thesaurus generation	Search engine	No	No	Thesaurus improves retrieval technologies in specific data domains.
[14]	Neural network	Free text database search	No	No	Unsupervised learning techniques address information retrieval challenges.
[15]	Cosine similarity	Education	No	No	Summarization techniques improve educational technology systems.
[16]	LSTM	Question answering	No	No	Big data analysis enhances development of question-answering systems.
[17]	TextRank algorithm	Search engine	No	No	TextRank-based algorithms improve information retrieval.
[18]	NLP	Legal	No	No	Expert-annotated NLP enhances AI-enabled information retrieval in legal services.
[19]	LSTM	Not specified	No	No	Proposed framework aims at robust and scalable information retrieval in big data contexts.
[20]	Document level similarity	Legal	No	No	Document similarity and pooling techniques enhance the quality of relevance judgments in information retrieval systems.
[21]	LSTM	General	No	No	Integrating textual information improves content-based image retrieval.
[22]	Deep-LSTM	Question answering	No	No	Good information retrieval using deep LSTM

Overall, these papers collectively underscore the significant advancements made possible through the application of deep learning and NLP techniques in improving information retrieval across various domains, although none specifically focused on retrieval or working with tables.

Having said that, the review identifies significant research gaps, especially in processing and retrieving images and tables alongside text summarization. None of the studies addresses the integration of multimedia content, which is crucial for comprehensive information retrieval systems. Current approaches focus mainly on text, neglecting the effective summarization and integration of visual data [23], [24]. Furthermore, while some papers mention educational technology, there is a lack of focus on enhancing student learning specifically. Research should explore adaptive learning systems that incorporate multimedia elements to create engaging and effective educational experiences. Personalizing content based on individual student progress and preferences is also underexplored [25]. Advanced deep learning models that handle multimodal data, combining NLP with computer vision, are needed but underrepresented in the literature.

There is also a shortage of practical implementations, real-world case studies, and research into user interaction and experience with multimedia retrieval systems [26], [27].

So, it has to be said that despite advancements in text-based retrieval, there is a clear gap in integrating multimedia elements and enhancing student learning. Addressing these gaps could lead to more comprehensive and effective information retrieval systems.

4. CONCLUSION

In synthesizing the expansive insights garnered from the detailed literature survey, it becomes unequivocally clear that the domain of information retrieval is at a critical juncture, characterized by a dynamic interplay of technological innovations, methodological advancements, and interdisciplinary collaborations. Central to this evolution is the pervasive integration and application of deep learning frameworks, which have substantially augmented the precision, scalability, and adaptability of contemporary retrieval systems. These advancements, however, are not without their complexities and challenges.

The review illuminates several nuanced dimensions, including the intrinsic algorithmic biases that can influence retrieval outcomes, the domain-specific intricacies that necessitate tailored methodologies, and the overarching imperative of scalability in an era of burgeoning data volumes. Furthermore, the paper underscores the transformative potential of interdisciplinary synergies, wherein the convergence of diverse domains and expertise catalyzes novel insights and innovative solutions. The paper highlights the fact that many research works focus on standard datasets and do not work on the real world data. Retrieval or working with the tables and images present in the original data is an important source of information, but none of the information retrieval papers identify this and work on it. So, in conclusion, there is a requirement for a system that works with this important information.

Collectively, the comprehensive exploration offered by the review of the literature elucidates the multifaceted nature of modern information retrieval, highlighting the remarkable strides made, the challenges yet to be surmounted, and the promising avenues for future research and innovation in this dynamic and pivotal domain.

ACKNOWLEDGMENTS

The authors sincerely appreciate the support provided by the management of Thakur College of Engineering and Technology.

FUNDING INFORMATION

The authors state no funding is involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sanket D. Patil	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓
Zahir Aalam	✓	✓	✓		✓	✓		✓		✓	✓	✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available at http://doi.org/ mentioned in all the references.

REFERENCES

[1] T. P. Nagarhalli, A. M. Save, and N. M. Shekokar, *Fundamental models in machine learning and deep learning*. Design of Intelligent applications using machine learning and deep learning techniques, 2021. [Online]. Available: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003133681-2/fundamental-models-machine-learning-deep-learning-tatwadarshi-nagarhalli-ashwini-save-narendra-shekokar

[2] "What is information retrieval?" Accessed: Dec. 29, 2023. [Online]. Available: https://www.geeksforgeeks.org/what-is-information-retrieval/

[3] S. Prakash, H. R. Shashidhara, and G. T. Raju, "The role of an information retrieval in the current era of vast computer science stream," *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, no. 3, p. 139, 2013, [Online]. Available: https://www.researchgate.net/publication/305209082

[4] O. A. Abass and O. A. Arowolo, "Information retrieval models, techniques and applications," *International Research Journal of Advanced Engineering and Science*, vol. 2, no. 2, pp. 197-202, 2017.

[5] "Issues in information retrieval," Accessed: Dec. 29, 2023. [Online]. Available: https://www.geeksforgeeks.org/what-is-information-retrieval/

[6] K. Subrmanian, G. Thangarasu, Z. Yanyan, and K. N. Kannan, "Deep learning based algorithm for efficient information retrieval in blockchain transactions," in *2024 6th IEEE Symposium on Computers and Informatics, ISCI 2024*, IEEE, Aug. 2024, pp. 264-269. doi: 10.1109/ISCI62787.2024.10667712.

[7] S. Mengmeng, L. Zhibin, W. Qingwei, H. Man, and X. Feiyang, "An effective retrieval method to improve RAG performance," in *Proceedings - 2024 7th International Conference on Data Science and Information Technology, DSIT 2024*, IEEE, Dec. 2024, pp. 1–5. doi: 10.1109/DSIT61374.2024.10881380.

[8] L. Stanescu and B. Savu, "Automatic assessment of narrative answers using information retrieval techniques," in *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems, FedCSIS 2019*, Sep. 2019, pp. 355-358. doi: 10.15439/2019F96.

[9] V. A. Mishra, M. Srivastava, M. S. Nithya, S. Mishra, S. Kannimuthu, and N. Divya, "Evaluating automated information retrieval techniques for improved data accessibility," in *Proceedings - 2023 IEEE International Conference on Paradigm Shift in Information Technologies with Innovative Applications in Global Scenario, ICPSITIAGS 2023*, IEEE, Dec. 2023, pp. 163-167. doi: 10.1109/ICPSITIAGS59213.2023.10527529.

[10] A. Pandey and A. Bhat, "A review on textual question answering with information retrieval and deep learning aspect," in *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023*, 2023, pp. 224-229. doi: 10.1109/ICICCS56967.2023.10142729

[11] H. N. Van, D. Nguyen, P. M. Nguyen, and M. Le Nguyen, "Miko Team: deep learning approach for legal question answering in ALQAC 2022," in *Proceedings - International Conference on Knowledge and Systems Engineering, KSE*, IEEE, Oct. 2022, pp. 1–5. doi: 10.1109/KSE56063.2022.9953780.

[12] S. Zhao, Z. Huang, H. Guo, X. Xiao, and R. Ye, "Deep learning-based retrieval algorithms for recommendation systems," in *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications, ICPECA 2023*, IEEE, Jan. 2023, pp. 1155-1160. doi: 10.1109/ICPECA56706.2023.10076102.

[13] H. Matsiuk, L. Dzhydzhora, N. Kunanets, and O. Matsiuk, "Application of information retrieval thesaurus for the improvement of information retrieval technologies in specific data domain," in *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, IEEE, Sep. 2021, pp. 437-440. doi: 10.1109/CSIT52700.2021.9648729.

[14] J. C. Scholtes, "Unsupervised learning and the information retrieval problem," in *1991 IEEE International Joint Conference on Neural Networks*, IEEE, 1992, pp. 95-100. doi: 10.1109/ijcnn.1991.170387.

[15] I. Benedetto, L. Canale, L. Farinetti, L. Cagliero, and M. La Quatra, "Leveraging summarization techniques in educational technology systems," in *Proceedings - 2022 IEEE 46th Annual Computers, Software, and Applications Conference, COMPSAC 2022*, IEEE, Jun. 2022, pp. 415-416. doi: 10.1109/COMPSAC54236.2022.00068.

[16] D. Aksonov, A. Gozhyj, I. Kalinina, and V. Vysotska, "Question-answering systems development based on big data analysis," in *International Scientific and Technical Conference on Computer Sciences and Information Technologies*, IEEE, Sep. 2021, pp. 113-118. doi: 10.1109/CSIT52700.2021.9648631.

[17] C. Xu, "Research on information retrieval algorithm based on textRank," in *Proceedings - 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation, YAC 2019*, IEEE, Jun. 2019, pp. 180–183. doi: 10.1109/YAC.2019.8787615.

[18] K. Kanhaiya, Naveen, A. K. Sharma, K. Gautam, and P. S. Rathore, "AI enabled- information retrival engine (AI-IRE) in legal services: an expert-annotated NLP for legal judgements," in *Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023*, IEEE, Aug. 2023, pp. 206-210. doi: 10.1109/ICAISS58487.2023.10250733.

[19] H. L. Nguyen, T. N. Trinh-Huynh, and K. H. Le, "Towards a robust and scalable information retrieval framework in big data context," in *Proceedings - 2022 9th NAFOSTED Conference on Information and Computer Science, NICS 2022*, IEEE, Oct. 2022, pp. 211-216. doi: 10.1109/NICS56915.2022.10013446.

[20] M. H. Joseph and S. D. Ravana, "Generation of high-quality relevant judgments through document similarity and document pooling for the evaluation of information retrieval systems," in *International Conference on Software, Knowledge Information, Industrial Management and Applications, SKIMA*, IEEE, Dec. 2022, pp. 261-265. doi: 10.1109/SKIMA57145.2022.10029459.

[21] Y. Srivarsha and V. M. Manikandan, "A novel content-based image retrieval scheme based on textual information," in *2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021*, IEEE, Jul. 2021, pp. 1-4. doi: 10.1109/ICCCNT51525.2021.9579478.

[22] D. V. Vekariya and N. R. Limbasiya, "A novel approach for semantic similarity measurement for high quality answer selection in question answering using deep learning methods," in *2020 6th International Conference on Advanced Computing and Communication Systems, ICACCS 2020*, IEEE, Mar. 2020, pp. 518-522. doi: 10.1109/ICACCS48705.2020.9074471.

[23] A. Aarab, A. Oussous, and M. Saddoune, "Optimizing arabic information retrieval: a comprehensive evaluation of preprocessing techniques," in *2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC)*, IEEE, May 2024, pp. 1-4. doi: 10.1109/ISIVC61350.2024.10577827.

[24] D. Mahapatra, C. Maharana, S. P. Panda, J. P. Mohanty, A. Talib, and A. Mangaraj, "A fuzzy-cluster based semantic information retrieval system," in *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, IEEE, Mar. 2020, pp. 675-678. doi: 10.1109/ICCMC48092.2020.ICCMC-000125.

[25] Y. Peng and Y. Yang, "Learning to rank for biomedical information retrieval," in *Proceedings - 2021 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2021*, IEEE, Mar. 2021, pp. 562-566. doi: 10.1109/ICITBS53129.2021.00143.

[26] R. R. Deshmukh and A. B. Raut, "Ontology based semantic information retrieval system using data ranking," in *International Conference on Self Sustainable Artificial Intelligence Systems, ICSSAS 2023 - Proceedings*, IEEE, Oct. 2023, pp. 1032-1038. doi: 10.1109/ICSSAS57918.2023.10331873.
[27] D. Kumar, D. S. Yadav, and D. Yadav, "Wavelet tree approach for searching and indexing in geographical information retrieval," in *1st International Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies, CE2CT* 2025, IEEE, Feb. 2025, pp. 58-61. doi: 10.1109/CE2CT64011.2025.10939677.

BIOGRAPHIES OF AUTHORS

Sanket D. Patil is an Assistant Professor and Training and Placement Officer at Vidyavardhini's College of Engineering and Technology. He holds a master's degree in computer engineering from the University of Mumbai, and has over 10 years of experience in both teaching and industry. Sanket has published several papers in reputed journals and conferences, focusing on data preprocessing and sentiment analysis. His industry experience includes serving as a Chief Technical Analyst at Tech5 Solutions, where he specialized in system design and project management. He can be contacted at email: sanketpatil7@gmail.com.

Dr. Zahir Aalam currently serving as the dean of training and placement, holds a Ph.D. in Engineering. With over 19 years of teaching experience, he has established expertise in the fields of embedded systems, wireless networks, machine learning, deep learning and sensor networks. His extensive background in both academia and specialized technical areas has been instrumental in shaping his leadership role in training and placement activities. He can be contacted at email: zahir.aalam@thakureducation.org.